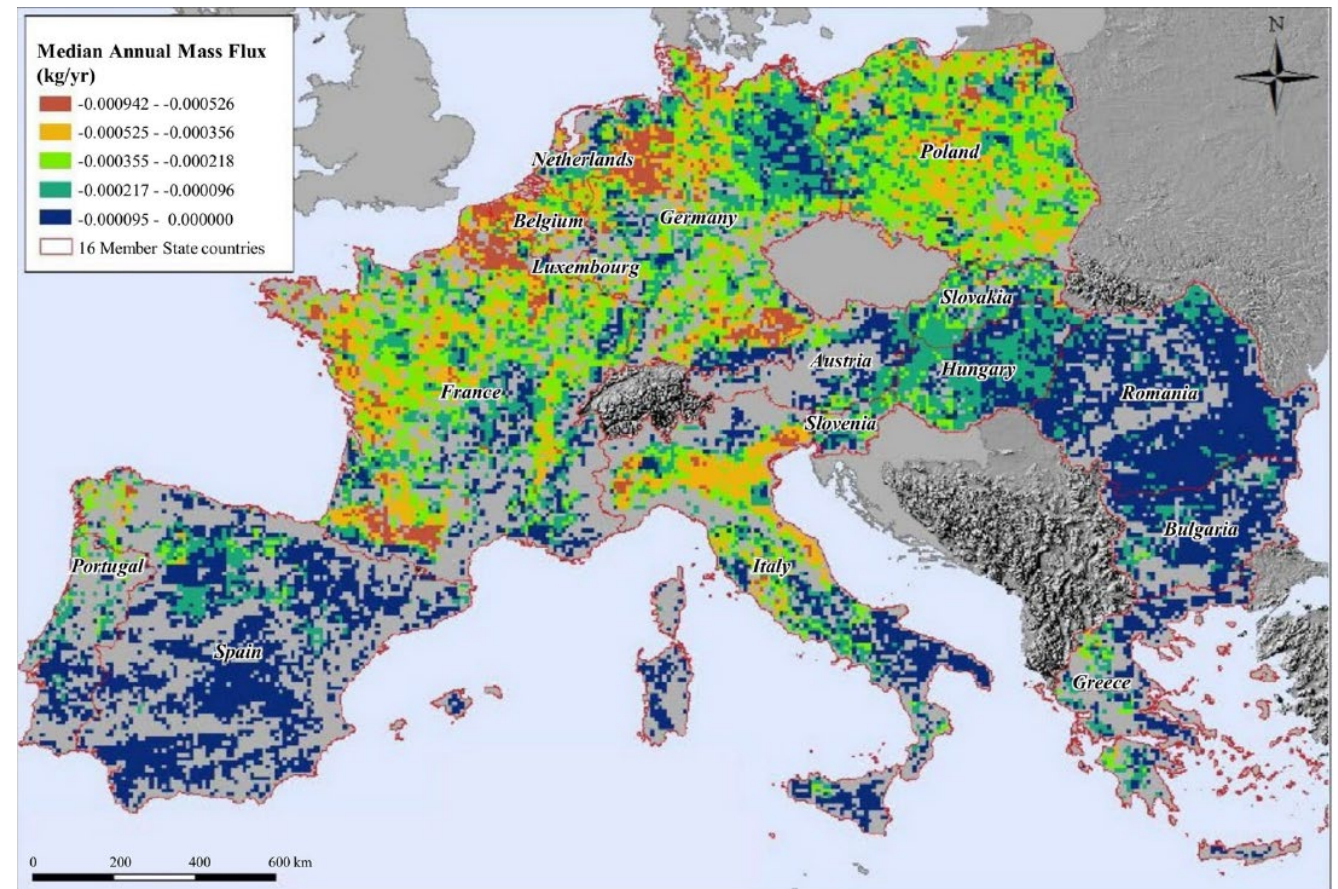


Version control for geodata storage and processing

Bernhard Jene



Version control for geodata storage and processing - why

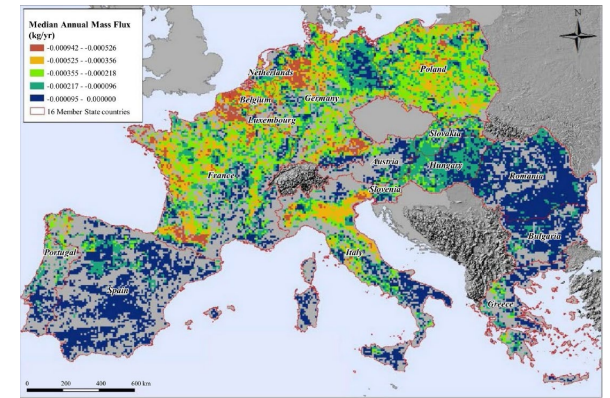
Motivation to bring it up

- Competing data sets of different quality existing
- Lack of information and quality control
=> Uncertainty of users which data set to use for which purpose
- Helps regulators to evaluate work that is done with a certain data set if it is under VC
- Was discussed and included in the resolutions in the European Modelling Workshop in October 2018 in Copenhagen (<http://pfmodels.org/emw9.html>):

“Version control is needed for high-resolution spatial databases for the EU”

- Was discussed and included in the Workshop about Spatially Distributed Leaching Modelling (SDLM) in May 2019 in Ghent (<http://esdac/jrc.ec.Europa.eu/sdlm>):

“Version control similar to that for the FOCUS software packages is necessary”



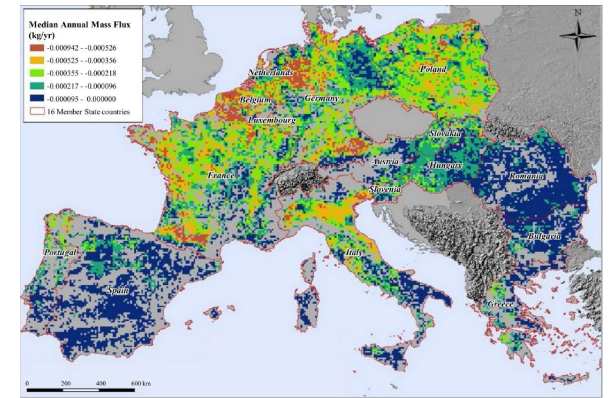
Version control for geodata storage and processing – what is needed?

1. Identification of geodata set

- Unique name
- Unique version number
- Release date
- Origin (institute), authors, sponsor

2. Clear documentation

- Good description of the data set:
 - What data are included (when, where)
- Spatial and temporal resolution
 - Measurement frequency
- Underlying raw data
 - Measurement method
 - Where is expert judgement involved
 - Homogeneity and consistence
- Derived data
 - Based on which raw data
 - Inter/Extrapolation applied?
 - Approach/model used (e.g. PTF, ET_p model) => references



Version control for geodata storage and processing – what is needed?

3. Quality control

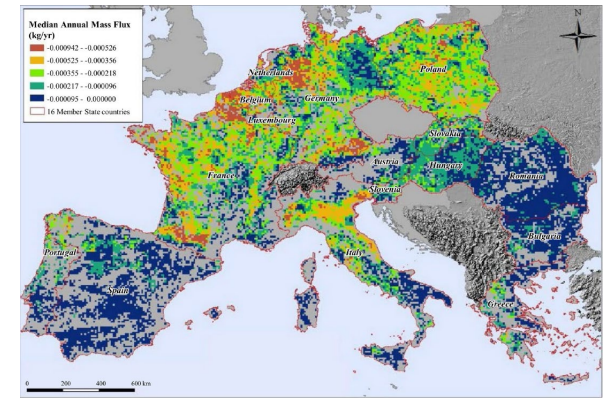
- What has changed to earlier versions
- Plausibility checks. Which? How many?
- Consistency checks. Which? How many?
- Comparison to other existing data sets with comparable data
- Check of approach/model used for derived data => still state of the art?

4. What can the data be used for

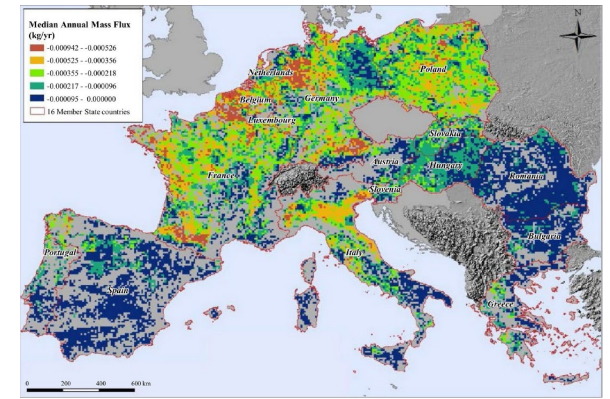
- Description of existing important applications
 - References if possible
 - Cannot be complete => not everything known
- Original purpose of the authors
 - Reason why the data were collected
- Possible new areas of application

5. Availability

- From where/whom can the data be obtained/downloaded
- Possibility for public access => one central place for geodata under VC (as for FOCUS models and documents?)
- Registration of user for info about issues or new versions



Version control for geodata storage and processing – what is needed?



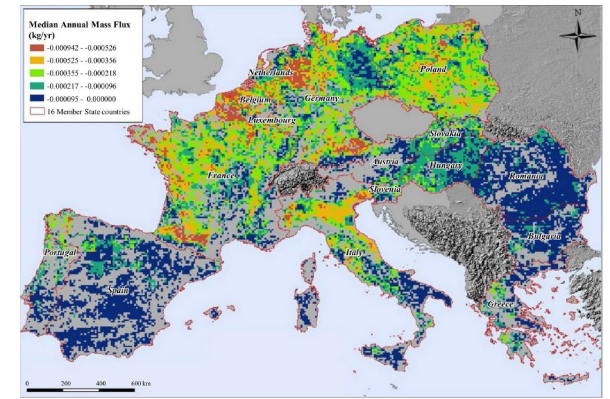
6. Costs

- VC includes additional effort => not for free => cost-benefit analysis?
- Who can do the work?
 - Data holders
 - Supporting institutions
 - Interested stakeholders

7. How to assure continuity

- Providing long-term capacity
 - References if possible
 - Cannot be complete => not everything known

Version control for geodata storage and processing – what is needed?



Conclusions

- Version control of geodata would be an important step forward
- Would be of advantage for all stakeholders
- Would improve quality, consistence and acceptance

Let's start